Primo real-time analytics/metrics

IGeLU 2015, Budapest, 3 September 2015
Session: 10.22 Room Q1: 15:30 – 16:00

Mehmet Celik
Senior Business Consultant
Who am I?

Mehmet Celik
Senior Business Consultant @ KULeuven/LIBIS

• Github:
  – http://github.com/mehmetc
  – http://github.com/libis

• Email: mehmet.celik@libis.be

• Twitter: @celikbe
LIBIS

- Business Unit @ KULeuven since the early ’70
- We provide library, museum and archives related services and consultancy see [http://www.libis.be](http://www.libis.be)
- We support Aleph, Primo, Alma, Rosetta, Omeka, Collective Access, …
- We host Aleph, Rosetta, Omeka, CA, …
- We are a Service Provider are hosted for Alma and Primo but for now we still have shell access
Service Provider

• We work closely with ExLibris
  – Need to know what is going on with the *hosted* environments. Status pages are not enough.
  – Need to provide quick solution to our customers

• ExLibris does extensive monitoring of all hosted environments.
  – Getting detailed information takes time. Tickets need to go up the chain.
We need our own monitoring

• Beginning this year we had a problem
  – Primo was slow, really slow
  – Finding the cause took a long time
  – It is not in our nature to sit back and relax

• Monitoring does not need a lot of data
  – ExLibris helped a lot
The idea

• Primo is written in Java and uses a lot of standard libraries
• Logging is taken care by -> Log4J
• Log4J can write to files but also to a network socket
  – We redirect parts of the logs to a locally hosted analytics environment.
Tools

- Primo jaguar.log
- [E]lastic search
- [L]ogstash
- [K]ibana
- Collector (CEP)

- ExLibris uses the same ELK environment for their daily monitoring
Data Flow

1. PRIMO FRONTENDS
2. Logstash
3. logstash_collector
4. Elasticsearch
5. Kibana
jaguar.log

- 2015-06-09 11:05:51,990 INFO [t-RMI TCP Connection(20)-31.186.254.54] [c-JaguarParallelMSearcher] - JaguarParallelMSearcher.search() start, for query: Q: ((wood)) AND (scope:(LUCA_LEMM_LIB)):KUL:false:400:L::dut
- 2015-06-09 11:05:52,025 INFO [t-RMI TCP Connection(20)-31.186.254.54] [c-JaguarParallelMSearcher] - JaguarParallelMSearcher.search() start, for query: SQ: ((wood)) AND (scope:(LUCA_LEMM_LIB)):KUL:false:400:L::dut
- 2015-06-09 11:05:52,099 INFO [t-RMI TCP Connection(20)-31.186.254.54] [c-JaguarSearchBean] - Elapsed Jaguar Search time: 116 milliseconds
- 2015-06-09 11:05:52,077 INFO [t-RMI TCP Connection(20)-31.186.254.54] [c-Searcher] - Found 2 document(s) (in 93 milliseconds) that matched query '((wood)) AND (scope:(LUCA_LEMM_LIB)):KUL:false:400' [1,10]
{
   "_index": "logstash-2015.06.22",
   "_type": "log",
   "_id": "AU4Z0tXe10JOufI5MV8j",
   "_score": null,
   "_source": {
      "@timestamp": "2015-06-22T05:52:00.584Z",
      "tid": "production-1-23326",
      "remote_ip": "10.24.0.103",
      "query": "((wood))",
      "frontend": 1,
      "system": "production",
      "query_count_local": 2167,
      "S": 48,
      "S2": 0,
      "D": 72,
      "D2": 0,
      "D3": 0,
      "F": 17,
      "DB": 28,
      "query_time": 227,
      "view": "PRIMO",
      "logged_in": "false",
      "language": "eng",
      "scope": "(scope:(32LIBIS_ALMA_DS))",
      "@version": "1",
      "type": "log"
   }
}
Visualizing the
What we noticed

<table>
<thead>
<tr>
<th>Time</th>
<th>query</th>
<th>query_time</th>
<th>DB</th>
<th>D</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 8th 2015, 06:02:06.719</td>
<td>((trends in cognitive sciences))</td>
<td>12347</td>
<td>1876</td>
<td>15</td>
<td>9392</td>
</tr>
<tr>
<td>March 8th 2015, 03:05:14.035</td>
<td>((Women's emotional adjustment to IVF a systematic review of 25 years of research.))</td>
<td>16179</td>
<td>0</td>
<td>2098</td>
<td>2787</td>
</tr>
<tr>
<td>March 8th 2015, 02:08:03.864</td>
<td>((lev manovich))</td>
<td>10509</td>
<td>2640</td>
<td>245</td>
<td>7115</td>
</tr>
<tr>
<td>March 8th 2015, 02:06:23.129</td>
<td>((new media a critical introduction))</td>
<td>20998</td>
<td>10405</td>
<td>19</td>
<td>10000</td>
</tr>
</tbody>
</table>
But …

• You need shell access to harvest and create these dashboards
The alternative … jQuery.PRIMO

- Client side convenience [library](#) for Primo
  - Free
  - MIT licenced
- Just a couple of lines to log searches as events in ELK
- [Gist](#) with configuration files
What is logged?

```javascript
{
    session_id:     jQuery.PRIMO.session.sessionId,
    ip:             jQuery.PRIMO.session.ip.address,
    fe:             jQuery.PRIMO.session.view.frontEndID,
    loggedin:       jQuery.PRIMO.session.user.isLoggedIn(),
    oncampus:       jQuery.PRIMO.session.user.isOnCampus(),
    count:          jQuery.PRIMO.query.count,
    scope:          jQuery.PRIMO.query.scope,
    sorted_by:      jQuery.PRIMO.query.sorted_by,
    step:           jQuery.PRIMO.query.step,
    page:           jQuery.PRIMO.query.page,
    tab:            jQuery.PRIMO.query.tab,
    search_type:    jQuery.PRIMO.query.type,
    text_query:     '',
    query:          {},
    facets:         {},
    view:           jQuery.PRIMO.session.view.code,
    ui_lang:        jQuery.PRIMO.session.view.interfaceLanguage,
    full_display:   jQuery.PRIMO.session.view.isFullDisplay(),
    ranking_categories: jQuery.PRIMO.session.user.ranking.categories,
    ranking_prefer_new: jQuery.PRIMO.session.user.ranking.prefer_new,
    deeplink:       jQuery.PRIMO.query.isDeeplinkSearch()
}
Raw data in Kibana
Demo
Conclusion

• ExLibris helped us a lot.
• Data can be gather from multiple environments -> different context = different focus.
• jQuery.PRIMO lets you create any data structure to log.
• Setting up ELK and sending data to it is easy.
Thank you!
Questions?